基于系统性综述的档案部门参与数据管理角色定位研究*

何思源 [1,2]

[1] 中国人民大学信息资源管理学院　北京　100872　[2] 中国人民大学电子文件管理研究中心

北京　100872

摘要：[目的/意义]探究档案部门参与数据管理的角色定位及路径方法，为档案部门在数据社会更好地履行档案管理职能、贡献档案力量提供理论支撑。[方法/过程]采用系统性文献综述分析国内外 104 篇档案参与数据管理相关文献。[结果/结论]档案部门参与数据管理存在接收者、指导者和协同者三种角色定位。档案部门作为接收者，通过数据归档联结数据生命周期和档案生命周期；作为指导者，运用来源原则、价值鉴定、连续性管理、长期保存等档案理论与方法及分类方案、目录等档案管控工具，为科学数据管理、开放政府数据、大数据应用等提供建议；作为协同者，从制度、主体、业务、资源、服务、系统等要素出发，与数据部门协同开展数据管理。后续研究需要在结构化数据归档对象识别与方案升级、档案部门作为指导者的实现机制及效果评价、档案与数据的业务协同及协同意愿与动机等方面进行更深入的理论探索与实证研究。

关键词：档案管理　数据管理　数据治理　数据档案化

分类号：G270

1 引言

数据及数据管理是政府、企业、学界等领域探讨的热点话题，加快培育数据要素市场、建立高效利用的数据要素资源体系上升到国家战略层面。一方面，数据管理使档案事业面临的生态环境、管理对象和工作内容发生巨大变革，业务数据化催生海量数据，档案管理对象空间走向“数据态”[1]，要求档案部门介入前端数据管理，确保对国家和社会有保存价值的数据符合档案管理要求。另一方面，档案部门作为信息机构的一员，有相对完善的规范体系、管理体系、工作流程和人员队伍，长期积累的档案化管理经验能够为建立健全数据管理体系提供支撑。档案部门通过参与数据管理，可以为数据社会贡献档案力量，推动档案工作提质增效，倒逼档案业务数字转型，提高档案事业在经济社会发展中的“贡献度”和“能见度”。

* 本文系中国人民大学 2021 年度拔尖创新人才培育资助计划成果。

作者简介：何思源，博士研究生，E-mail:2576620771@qq.com。

参与数据管理是档案部门释放价值、顺势而为的重要选择。在此背景下，我国的数字转型与数据治理相关规划多次提到档案及档案工作，将其作为数据社会发展的重要保障，体现出“档案参与”的特点[2]；与此相应，各地档案事业发展规划也积极倡导推动档案全面纳入大数据战略[3]。档案部门如何参与数据管理、如何在数据社会更好地履行职能成为亟待解决的现实问题。

目前，已有部分学者梳理了档案与数据相关领域的研究进展。例如，高晨翔等[4]将国内“档案数据”相关研究分为档案工作自动化阶段（1992—2003）、档案信息化时期（2004—2011）、大数据时代（2012—2019）三个阶段；金波等[5]从档案数据的概念、产生背景、管理对象和管理方法等角度梳理档案数据的研究现状；杨茜茜[6]从概念辨析、内容结构和功能定位等角度分析档案数据治理的研究进展；Rebecca Grant[7]聚焦科研领域，探讨数据、研究数据与文件，数据管护与研究数据管理技能，数据管理与文件管理路径，档案馆保管的研究数据等主题。但这些综述仍然存在一定局限，无法勾勒出档案参与数据管理相关研究的整体面貌，主要表现在 2 个方面：第一，就研究主题而言，多聚焦档案部门内档案数据的管理与治理，关注“引进来”（即数据管理理论与方法在档案领域的应用），较少涉及档案部门在数据管理中的角色定位。第二，就研究范围而言，相关中文综述较少涉及英文文献，而英文综述只关注科研领域，没有涉及更广泛的档案参与数据管理，且发表时间较早。在数据要素市场化的背景下，数据管理格局发生显著变化，亟需从理论层面系统梳理档案参与数据管理的理论成果，服务当下的档案与数据管理实践。

因此，本研究将采用系统性文献综述总结国内外档案参与数据管理的相关研究，批判分析当前研究的局限和不足，展望未来研究趋势。具体研究问题有 2 个：一是档案部门在数据管理活动中的角色定位有哪些；二是在不同的角色场景中，档案部门参与数据管理的路径方法有哪些。需要说明的是，在本研究中，“档案部门”是指档案馆、档案室、档案主管部门等的统称；“数据管理”是指“为了交付、控制、保护并提升数据和信息资产的价值，在其整个生命周期中制定计划、制度、规程和实践活动，并执行和监督的过程”[8]1，既包括宏观层面的数据治理，也包括微观层面数据生命周期各环节的具体工作。

2 研究方法

本研究采用系统性文献综述探究档案部门参与数据管理的角色定位及路径方法。首先是文献检索；其次，根据研究问题及检索得到的文献确定文献筛选标准（见表 1），并进行筛选和去重；最后，通过引文追踪及研究者日常积累进行文献补充，得到 104 篇文献，如图

1。需要说明的是，由于档案学规模较小，部分学科领域内期刊未被主流数据库收录，故对澳大利亚研究理事会（Australian Research Council）评出的 6 本档案学 A+和 A 类期刊[9]的官网单独进行检索，以提高查全率和查准率；考虑到档案学的 A+和 A 类期刊均未被 SSCI 索引，故未对英文期刊论文的范围作出限制。

第一步：文献收集

文献来源-数据库

中文：CNKI系列数据库、中国人民大学学位论文库
英文：WOS-core collection、Emerald、Springer、Taylor & Francis、ScienceDirect、Wiley Online Library

文献来源-期刊官网

Archival Science、*American Archivist*、*Archivaria*、*Archives and Records*、*Archives and Manucripts*、*Records Management Journal*

检索式

中文：档案/文件/归档+数据
英文-数据库：record/archive+data
英文-期刊官网：data

检索字段

数据库检索字段：题名
期刊官网检索字段：题名或主题

中文文献7270篇，英文文献8558篇，共计15828篇

第二步：文献筛选

第一轮筛选

标准：文献类型、学科领域、刊物声望
筛选对象：检索得到的15828篇
筛选范围：文献类型、学科、刊物名称

中文文献856篇，英文文献759篇，共计1615篇

第二轮筛选

标准：研究对象、研究内容
筛选对象：第一轮筛选后的1615篇
筛选范围：篇关摘、全文

中文文献55篇，英文文献43篇，共计98篇

第三步：文献去重

去重

中文：55-4=51篇
英文：43-6=37篇

中文文献51篇，英文文献37篇，共计88篇

第四步：文献补充

文献补充

策略：引文追踪、研究者日常积累
中文：51+3=54篇
英文：37+13=50篇

中文文献54篇，英文文献50篇，共计104篇

图 1 文献收集与筛选的流程与方法

表 1 文献筛选标准

维度	纳入标准	排除标准
文献类型	期刊论文、学位论文、专著	会议论文、书评、新闻报道等

研究对象	档案部门参与数据管理活动的相关研究	数据管理理论、方法与技术在档案领域应用的研究；档案数据化开发利用的研究；档案部门开展档案数据治理/管理的研究；档案元数据研究；名称中虽带有“档案/归档”字样但并非传统档案学意义上的相关研究（如网络存档研究等）
研究内容	方法科学、论证合理、观点清晰的研究	与档案理论和工作实际存在较大偏差的研究
学科领域	信息资源管理/图书情报与档案管理学科	计算机科学与技术、健康医疗等其他学科领域
刊物声望	中文期刊论文限定 CSSCI 和北大核心 英文期刊论文不作限定	中文非 CSSCI 和北大核心的期刊论文

最终获取的文献集的特征分布如表 2 所示。这些研究大多发表于近十年，与大数据发展趋势基本吻合；中英文占比基本持平；以期刊论文为主，近五年开始出现该主题的学位论文和研究专著。

表 2 文献集的特征分布

文献特征		文献数量/篇	占比/%
语言类型	中文	54	51.92
	英文	50	48.08
发表时间	2000 年以前	9	8.65
	2001—2010	12	11.54
	2011 年至今	83	79.81
文献类型	期刊论文	96	92.31
	学位论文	5	4.81
	专著	3	2.88

本研究采用编码方法进行分析。首先，逐篇阅读文献，对文献中出现的关于研究问题的核心观点进行标记、编码，并在分析过程中逐步完善，最终得到稳定的编码结果，编码示例如表 3 所示。

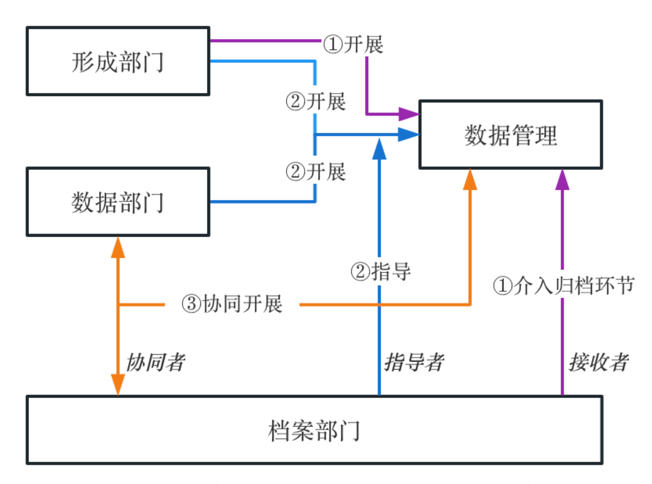
表 3 编码示例表

文献来源	观点原文	编码
John McDonald 等（2014）	Analysis of the business processes supporting big data and open data initiatives should form the basis for addressing the issue of retention and disposition.	接收者-数据归档范围-鉴定方法
	……	……

Charlotte Borgerud 等（2020）	Therefore, extended collaboration between researchers and archivists, where the archivists should play a consultative part in what data should be kept and for how long, is essential.	指导者-以价值鉴定指导数据挑选
	……	……
梁凯（2018）	档案管理部门在数据质量控制上与大数据管理部门存在着共识……	协同者-业务协同-数据质量管控
	……	……
……	……	……

3 研究结果

根据当前文献，可将档案部门参与数据管理的角色定位分为接收者、指导者和协同者，如图 2。在接收者场景中，数据形成部门对数据进行管理，档案部门介入归档环节，对数据进行接收和长期保存，强调数据资源从形成部门流向档案部门；在指导者场景中，形成部门或数据部门作为数据管理主体，档案部门对数据管理进行指导和协助，强调档案理论、方法与工具从档案领域走向数据领域；在协同者场景中，档案部门与数据部门协同开展数据管理。三种角色定位并不是非此即彼、相互割裂的关系，而是互为支撑、共存共生的关系。例如，档案部门作为接收者积累的理论与方法能够为提供指导和实现协同奠定基础，作为指导者和协同者形成的工作成果也能为提高归档数据质量提供保障。



注：①（紫色）代表接收者场景；②（蓝色）代表指导者场景；③（橙色）代表协同者场景。

图 2 档案部门参与数据管理的角色定位

3.1 作为接收者参与数据管理

3.1.1 基本逻辑

根据国际数据管理协会（DAMA International）等数据管理行业组织提供的理论模型，数据生命周期通常包括从规划设计到存储和维护、再到利用和增强并最终处置的完整生命周期[8]11。其中，“处置”包括销毁、归档保存等行为。根据《中华人民共和国档案法》，档案部门的核心职责是档案的“收集、整理、保管和提供利用”[10]。对国家和社会有保存价值的数据属于档案，需要归档[11]，因此，立足本职工作，基于现行档案管理模式，档案部门可作为接收者参与数据管理（如图 3）。在此场景中，档案部门通常被视作资源的保管者和守护者，符合人们对档案部门最传统和最经典的形象认知，档案部门通过数据归档架起数据生命周期和档案生命周期沟通的桥梁。需说明的是，在档案学文献中，尤其是在数据归档的语境中，“数据”通常是指以数据库等形式存在的结构化数据[12]，因此接下来将聚焦结构化数据归档的相关研究。

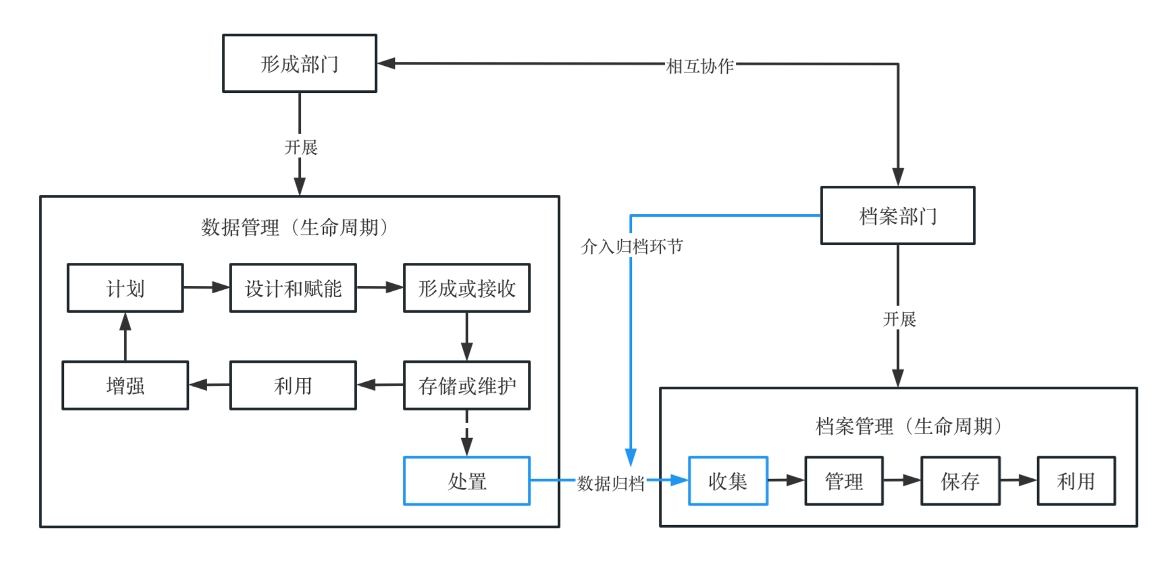


图 3 档案部门作为接收者参与数据管理的基本逻辑

3.1.2 归档范围

在数据归档过程中，档案部门首先要明确归档范围，清晰界定管理对象，聚焦具有档案价值和档案属性的数据[13]，因为档案管理的全部核心、逻辑起点和依赖路径是价值鉴定[14]。关于归档范围的讨论可分为两个层面：

首先，就单份档案而言，需要在理解数据结构的基础上，明确档案构成要素，建立识别归档对象、确定归档单元的方法论体系。一种观点倾向于从数据形成角度出发，将数据视作数据采集等数据形成业务活动的产物，从整体上将数据库、数据体作为归档单元进行鉴定、

归档[15]；另一种观点倾向于从数据利用角度出发，将数据视作查询访问等数据利用业务活动的产物，将呈现在用户面前的视图等数据对象作为归档单元[16]。以温哥华市地理信息系统为例，既可以将档案界定为整个地理信息数据库，也可以将档案界定为用户创建的特定地图视图[17]。有学者综合上述两种视角，提出更为灵活的“档案数据体”概念，即因系统管理需要而经档案化组织聚集在一起的数据集合[18]，在物理层面（数据形成视角）是易于系统管理的数据，在逻辑层面（数据利用视角）是可识别的档案。

其次，就档案集合而言，需要明确哪些数据具有归档保存价值。大数据时代，数据存储空间持续增长且成本降低，数据检索利用能力不断提升，有学者提出弱化档案价值鉴定[19]和数据全量留存[12, 20-21]的观点，也有学者提出异议[22-23]，激起学界广泛讨论[24]。由于计算能力、存储空间和管理资源等现实因素的限制，在可预见的未来，数据全量留存并不现实，但可以适度放宽归档范围，保存更多的数据[24-25]。就鉴定方法论而言，一种观点是从职能角度出发，采用基于职能的业务流程分析方法制定数据归档范围和保管期限表[26]；一种是从信息系统出发，按照信息系统梳理所有数据，制定归档范围和保管期限表[27]，基于此，中石油尝试将信息系统分为生产运行类、连续监测类、经营管理类和综合管理类，每类系统有不同类型的数据，对应不同的归档范围[28]。综上，档案部门可以探索“职能—系统—数据”三要素联动的归档范围识别方法[18]。

3.1.3 归档思路

数据归档思路主要有物理归档和逻辑归档两种。其中，物理归档是“根本之道”，逻辑归档是新技术环境下的妥协性选择、过渡性方案[29]。目前，具有实际可操作性的成果集中在物理归档，主要有技术典藏、留存快照、降维归档、数据交换等 4 种思路（见表 4）。此外，也有学者提出基于数据仓库的数据库归档思路[15]，但数据仓库更多地面向决策支持，而非归档保存，且数据仓库构建逻辑与档案价值鉴定标准存在显著差异。对结构化数据归档而言，重要的不是保存数据的外观，而是维护数据再现的能力。因此，无论采用何种思路，都需要考虑“语义保存”的问题，重点关注规则、模型、算法等数据所依赖的语义逻辑[30]，在保存数据本身的同时保存数据的相关文档（如布局、代码、算法、功能、用户手册等），并用元数据及说明文档对数据进行描述[30-34]，确保数据能在未来按照既定的逻辑再现。

表 4 结构化数据物理归档的方法

归档方法	基本含义	相关评价	典型案例
技术典藏	保存数据库文件及其所依赖的	成本高，无法根本解决软硬件	理论上存在，实践中未

	软硬件系统	过时带来的可用性问题，不具备现实可操作性，无法实现归档后的有效利用[16,25]	广泛采用
留存快照	保存某个特定时间点关于指定数据集合的完全可用拷贝，快照频率取决于数据修改和删除的频率、法律或业务要求及数据库逻辑结构的修改频率等[33]	主要适用于静态数据库，不适用于动态实时更新的数据库，会丢失两次快照时间间隔内形成或发生变化的数据[33]	英国国家数据集数字档案馆[33]、温哥华市地理信息系统 VanMap[17]、苏州工业园区地理信息库[35]
降维归档	将结构化数据“降维”到半结构化数据（如 XML）进行归档保存	在解决系统依赖性、提高可理解性方面有一定优势，但在维护数据完整性及数据利用等方面存在劣势[28]	瑞士联邦档案馆的 SIARD 及其在美国、韩国、澳大利亚等国的应用[36-39]，我国的《档案关系型数据库转换为 XML 文件的技术规范》（DA/T 57—2014）和《青岛市机关业务信息数据库归档与移交办法（试行）》等
	将结构化数据“降维”到非结构化数据（如 PDF）进行归档保存		在国内占主导地位，如中石油[28]、中石化[40]等
数据交换	采用数据交换标准进行归档，将结构化数据转为中间格式（如在三维数据中广泛采用的 STEP）	转换过程会造成数据丢失，档案工作者需要清楚哪些信息可丢失，哪些不能丢失[41]；对丢失且有价值的信息，可在中间格式基础上，基于元数据等技术，使其承载更多信息[25]	在三维数据归档领域占主流，如空客德国产品数据归档[41]

对档案界而言，逻辑归档并不陌生，但当前研究主要关注逻辑归档的理念与构想及物理归档与逻辑归档之争等内容，较少涉及逻辑归档的实现机制、路径、方法、要求等内容。若逻辑归档，档案部门需要接收并保管数据的目录[31,42]及部分元数据[29]，并将业务凭证和有机关联等档案专业追求[18,43]融入数据的实体管理活动中。以电子证照为例，在形成活动中，需保存主体、职能、事由及版本等方面的元数据，以维持来源联系和版本联系；在利用活动中，证照利用者需保存资源内容、形成背景等元数据，提供者需保存利用主体、缘由等元数据[44]。

3.2 作为指导者参与数据管理

3.2.1 基本逻辑

后保管时代，档案部门的角色不仅是保管者，也是指导者。作为指导者的档案部门应立足档案领域基本理论与方法，为解决数据管理问题提供指导和建议。从数据管理场景的角度来看，科学数据管理受到的关注最高，其次是政府开放数据和大数据应用；从档案理论、方法与工具的角度来看，可为数据管理提供支持的档案理论与方法主要有来源原则、价值鉴定、连续性管理（包括文件生命周期理论、文件连续体理论）、长期保存等，可为数据管理提供支持的档案管控工具有分类方案、案卷登记册及目录等检索工具，其中，来源原则是核心和根基，因为对来源与背景的理解与反思是档案工作者最珍视的价值观[45]，如表 5 所示。

表 5 可支持数据管理的档案理论、方法与工具

数据管理场景	文献依据	可支持数据管理的档案理论、方法与工具
科学数据管理	Micah Altman 等（2009）	价值鉴定[46]
	Christopher A. Lee 等（2011）	在集合层面进行管理、背景信息、记录来源、与形成者沟通、关注预期的二次利用需求[45]
	Dharma Akmon 等（2011）	挑选、保存背景信息、生命周期管理（前端控制）[47]
	Kathleen Fear 等（2012）	来源原则[48]
	Kristen A. Yarmey 等（2013）	价值鉴定（侧重数据质量评估）、元数据[49]
	Daniel Noonan 等（2014）	价值鉴定、信息组织（如著录等）、长期保存[50]
	Joanne Evans 等（2014）	文件管理信息学（包括连续体思想、文件管理元数据、信息文化分析、业务流程分析、访问获取分析）[51]
	Sue Childs 等（2014）	价值鉴定、敏感性审查[52]
	Alex H. Poole（2015）	来源、挑选与鉴定、真实性、元数据、风险管理、信任[53]
	Rebecca D. Frank（2015）	来源原则[54]
	毛天宇（2016）	来源原则、档案价值鉴定理论、文件生命周期理论[55]
	Charlotte Borgerud 等（2020）	生命周期管理（前端控制）、元数据、价值鉴定[56]
	王宁等（2021）	连续性保障、背景信息管理、鉴定处置、长期保存[57]
	于英香等（2021）	来源原则、鉴定理论、文件生命周期理论、文件连续体理论[58]
开放政府数据	Julie McLeod（2012）	长期保存、生命周期管理（前端控制）、鉴定处置[59]
	Lluís Esteve Casellas Serra（2014）	分类方案、案例案卷总登记册、价值鉴定、来源原则、长期保存[60]

	Valerie Léveillé 等（2015）	分类方案[61]
	周文泓（2018）	价值鉴定、长期保存、生命周期管理（全程管理）、数据质量管控[62]
	Elizabeth Shepherd 等（2019）	来源、职能分类、元数据、风险管理、平衡隐私与公共利益、规避不当发布数据带来的风险、确保数据准确[63]
大数据应用	Victoria Louise Lemieux 等（2014）	以 ISO 15489 和 ARMA 原则为代表的文件管理方法论[64]
	Kevin L. Dale（2015）	处置、统一数据目录、生命周期管理（全程管理）[65]
	Thomas Sødring 等（2020）	隐私保护[66]
一般意义上的数据管理（未明确限定场景）	周林兴等（2021）	文件连续体理论[67]
	刘越男（2021）	凭证性保护（如文件对象构成、元数据、四性保障与检测等）、价值鉴定、信息组织（如维护有机关联、元数据方案等）[29]
	章燕华（2022）	凭证性维护、群体性治理、集约化管理、社会化利用、集体性记忆构建、制度化体系建设[68]

围绕开放、共享与应用的数据管理核心目标，档案理论、方法与工具主要解决两类数据问题（如图 4）。一类是基础问题——数据范围把控，关注当前应留存哪些数据供未来利用，哪些数据可以开放共享。档案部门长期关注价值鉴定，在评估数据价值方面积累了丰富经验，且具有分类方案、数据目录等一系列管控工具，对数据范围把控具有借鉴意义。一类是核心问题——数据质量管控。采用来源原则，通过分类方案呈现有机联系，通过元数据捕获背景信息，能为数据提供更多的解释性信息，构建可信数据；采用连续性管理思维，通过前端控制和全程管理，将质量要求前置到数据形成乃至系统设计阶段，能实现数据质量源头治理；采用迁移、仿真、封装、格式登记等长期保存策略，能确保数据在未来持续可用。

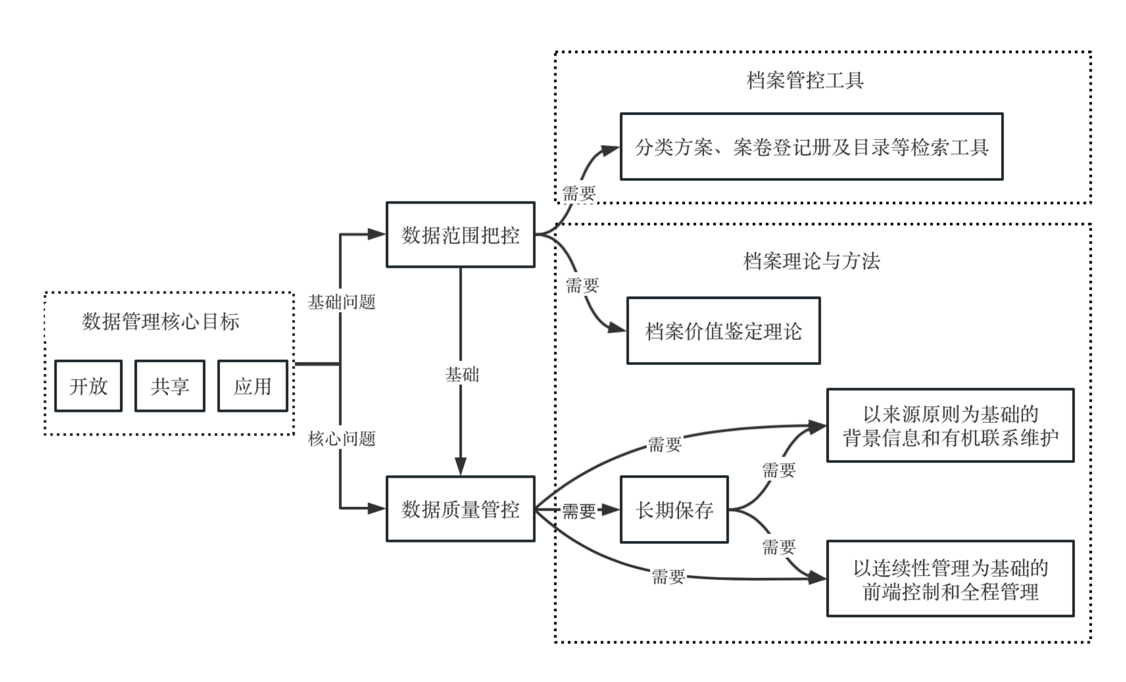


图 4 档案部门作为指导者参与数据管理的基本逻辑

3.2.2 档案部门作为指导者参与科学数据管理

首先，随着时间推移，科学数据的价值存在巨大差异[52]，考虑到成本收益和利用效率[57]，永久保存持续增长的所有科学数据并不现实，档案工作者可凭借档案价值鉴定专业知识，为科学数据价值评估提供基本框架[50, 52]。根据谢伦伯格（Schellenberg）的档案双重价值论，科研人员虽然清楚科学数据的第一价值，但较少考虑第二价值，档案工作者可为其提供关于第二价值的培训和指导[50]。在实践中，美国社会科学数据保存联盟（Data Preservation Alliance for the Social Sciences）就采用了档案价值鉴定标准和实践[46]。根据档案价值鉴定的来源标准、内容标准、形式特征标准、相对价值标准、效益标准等[50, 69]，可从数据来源和背景的重要性、收益（数据的科学价值、历史价值、经济价值及数据重用的可能性）、成本（包括处理成本、保存成本、替代成本等）、数据质量（准确可靠性、非冗余性、完整性、可用性、独特性等）、相对性（与保管机构任务的相关性、与馆藏数据的互补性）、法律道德限制等方面进行科学数据价值评估[46, 50, 69]。

其次，档案工作者可借助来源原则、连续性管理、长期保存等方面的经验提高科学数据质量，促进数据重用。来源原则要求维护科学数据的背景信息和有机联系。Kathleen Fear 等学者的实证研究表明，结合科学数据仓储中提供的其他信息及科研人员的经验和先验知识，来源信息及档案联（archival bond）能够帮助科研人员确定数据可信度[48]。一方面，需要充分记录数据的形成环境、处理过程、质量控制措施等背景信息[70]。这些信息需要通过高质量

元数据（通常表现为结构化形式）及科学数据说明文档（通常表现为非结构化文本形式）呈现[47,71]。另一方面，需要维护科学数据之间的有机联系，按照业务来源对科学数据进行分类整理[57]，同时也要维护科学数据与其他科研文件材料（如标本、出版物）之间的有机联系[48]。连续性管理要求提前谋划科学数据管理。由于数据生命周期每个阶段所采取的行动都具有累积效应，因此档案工作者需要理解数据生命周期，将鉴定、记录背景信息等实践活动前置到数据生命周期早期阶段[70]。尤其是在科研资助机构要求项目申请者提交数据管理计划（Data Management Plan，DMP）的背景下，档案工作者可将相关要求前置到项目申请阶段，通过 DMP 明确定义数据的生成方式、元数据、格式、存储、互操作性等[72]。长期保存要求数据形成者采取合适的存储和保存策略，档案工作者可以为科研人员提供存储库选择建议及数据管护最佳实践[50]。

遗憾的是，Charlotte Borgerud 等学者的研究表明，虽然档案工作者理应成为科学数据管理及开放科学数据的重要成员，但事实上，受制于自身知识结构、纸质档案管理传统、时间和精力等因素，相当一部分档案工作者并不愿意承担或参与科学数据管理活动[56]。

3.2.3 档案部门作为指导者参与开放政府数据

首先，在开放政府数据运动中，政府需要清楚自己掌握哪些数据，并有效检索这些数据[73]，档案部门可凭借其管理工具和经验辅助开放数据集的定位和挑选。基于分类方案和案例案卷总登记册，可从整体上把握组织机构的职能、活动、形成者、文件及构成文件的数据，形成数据地图，结合应用程序目录、个人数据案卷登记册等其他管理工具，依据价值鉴定的经验，解决组织机构有哪些数据、数据在哪、哪些数据可以开放、哪些敏感数据不能开放等问题[60-61]。

其次，可靠的高质量文件是开放政府数据的基础[73]，管理不善的文件将无法传递准确、可靠、完整的数据，从而带来数据的误解与滥用、掩盖欺诈、错误的政策建议和投资等[73-74]，档案理论与方法是确保开放数据可信的重要手段。从来源原则的角度来看，背景信息维护是确保数据可追溯、真实可靠的核心要素[73]。因此，关于开放数据集的描述必须准确、清晰、精确，档案领域的方法可提供关于数据集内容及来源、生成数据的法律依据、数据提取限制等方面的描述[60]，确保数据集可追溯至原始可靠的数据源[63]，实现数据可信。从连续性管理的角度来看，只有在数据形成时进行管控才能有效保障数据质量，如果事后管理，数据质量将变得复杂[60]。因此，档案工作者需考虑数据全生命周期，在全程管理框架下及时参与数据管理[62]，可以指导系统设计者和数据形成者通过良好的元数据和互操作的系统实现数据的

捕获和管理[59]。从长期保存的角度来看，档案工作者可以介入开放数据产品的管理和保存[60-61, 75]，为开放数据的长期归档保存提供建议，确保这些数据在未来仍然可用、可分析。

3.2.4 档案部门作为指导者参与大数据应用

大数据应用是释放数据价值、激发数据活力的重要环节，但在现实中面临着一系列挑战，尤其是数据预处理带来的困难，包括数据无法获取、数据碎片化、数据质量（如缺少值、格式问题、非标准化）、数据转换、数据形成与管理方式同用户利用需求脱节等[64]。Victoria Louise Lemieux 等学者通过对比分析发现，以 ISO 15489 和 ARMA 基本原则为代表的文件管理作为一种更基础、更广泛的信息治理方法论，能有效解决大数据应用中的上述问题[64]。

首先，档案工作者能够通过建立数据及数据存储库的统一目录，掌握组织机构的信息流及信息使用情况，实现数据可发现和可检索，确保数据分析人员清楚数据位置；也能通过处置行为清除重复或不必要的数据，通过控制措施减缓或阻止非受控数据存储库的创建，降低数据分析成本[65]。

其次，档案理论与方法可有效提高数据质量。从来源原则的角度来看，由于大数据可能不准确或蕴含潜在的不平等，如果缺乏元数据，这些负面影响将会在分析应用活动中随时间推移而加剧，档案工作者可借助专业知识为大数据提供充足的背景信息[76]。也就是说，档案工作者可为数据分析人员提供数据的背景、结构、语义等信息及保管链的历史，并通过背景元数据弥合数据形成和利用之间的脱节，增强数据分析人员对数据的理解和信任[64]。从连续性管理的角度来看，在数据形成阶段提供技术控制、组织控制和行为控制等质量控制方面[64]的建议，确保数据在全生命周期都得到组织和积极管理[65]。从长期保存的角度来看，可提供仿真、捕获文件快照、元数据著录等确保大数据长期可用的方式[76]。

3.3 作为协同者参与数据管理

3.3.1 基本逻辑

20 世纪 80 年代，就有学者率先提出档案管理与数据管理协同的理念[77]，但并未引起广泛关注，直至近些年，独立于档案管理体系的体系化数据管理兴起，档案管理与体系化数据管理的边界出现交汇和冲突，协同问题才渐渐得到重视。档案部门与数据部门协同开展数据管理的基本逻辑如图 5 所示。

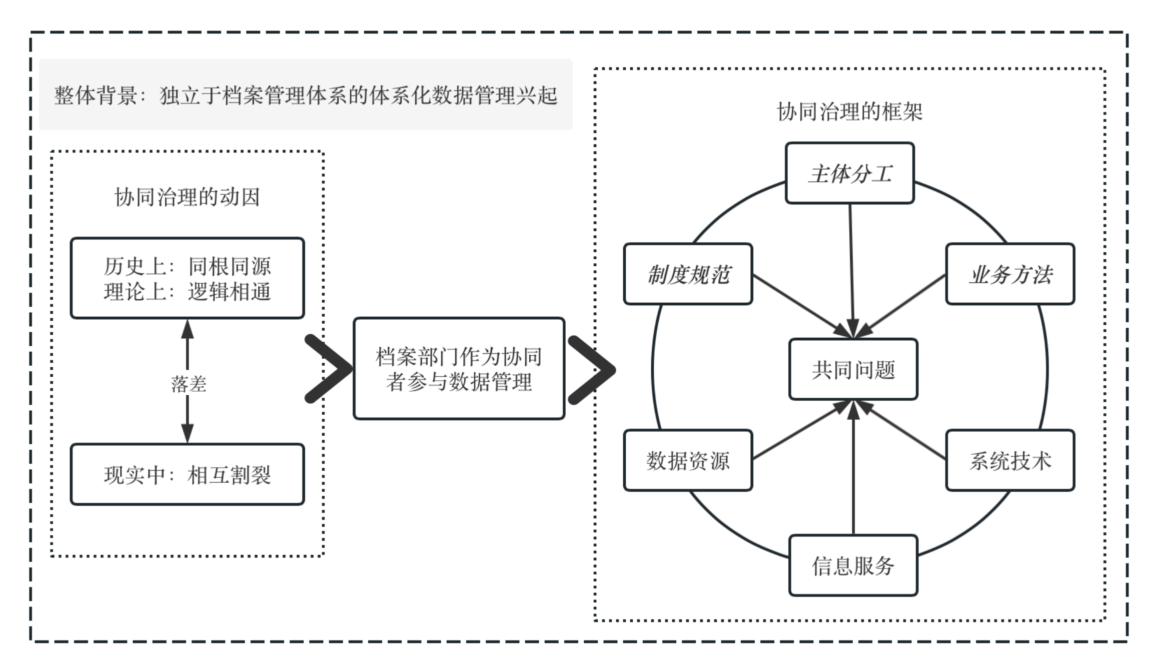


图 5 档案部门作为协同者参与数据管理的基本逻辑

从历史维度看，档案管理与数据管理同根同源，有着共同的历史烙印，可追溯至公元前 4000 年的苏美尔文明[78]。**从理论维度看**，档案管理与数据管理的核心概念与话语体系在底层逻辑上共通[49, 79]，具备协同基础与合作空间。虽然档案与数据的概念纷繁复杂、角度各异，但无论采用何种理解，二者始终存在紧密联系，或为包含关系[29, 80-81]，或为交叉关系[14, 82]，甚至可同义替换[77]。而且档案管理与数据管理在目标理念、管理流程、质量要求等方面[14, 45, 78, 82-84]也具有一致性。**从现实维度看**，档案管理和数据管理在体制、制度、业务、行业组织及从业人员的教育背景、技能、技术与工具等方面呈现平行发展和相对割裂的状态[29, 77, 82, 84-85]。这种现象带来一系列现实问题，包括业务要求冲突，尤其是在管理对象紧密关联的情况下，档案部门和数据部门提出的不同要求将加重数据形成者的负担[78, 83]；档案部门与数据部门的管理重叠与管理真空并存，信息资产存在流失风险[82]；资源重复投入与配置失衡，冲击较为弱势的档案管理体系[78, 82-83]；阻碍信息资源统一发现与共享利用[82]，因此，需要协同档案管理与数据管理。此外，档案部门和数据部门及档案管理与数据管理优势互补[78, 82, 86]、数字化转型背景下面临的共同难题[78]等也是推动协同的现实因素。

协同既有其历史与理论根基，也有其现实必然性，已成为档案界乃至信息资源管理领域的重要议题。有学者在档案与数据协同治理方面展开探索性、框架性研究，提出囊括共同问题、制度、主体、业务、资源、服务、系统等内容的协同框架[84, 87]，据此系统考察了国内外科研档案和科学数据协同治理的实践现状[88]，明确了档案部门作为协同者参与数据管理的基本要素。由于现有文献主要关注主体、制度、业务等方面的协同，接下来主要围绕这三个方面作进一步的分析。

3.3.2 主体协同

主体协同关注责任主体间协同关系的建立与维护。在实际协同过程中，受制于认知、体制、制度、能力等因素，档案部门与数据部门在协同意愿、职能关系、沟通联系、协同内容等方面存在问题[89]。为解决这些问题，首先，需要建立协同治理组织架构。理想状态是建立信息资源（包括档案和数据）综合治理机制[88, 90]，将档案、数据、业务、信息化等部门纳入其中[91]，但更现实的方式是实现“双向纳入”，即将档案部门纳入数据议事协调机制[90, 92]，将数据部门纳入档案议事协调机制[92-93]。其次，需要根据协同治理组织架构等因素灵活选择一体式、委托式、业务集成式、资源交换式等具体协作模式[87]。再次，利益主体间不可避免地存在一系列矛盾冲突[13]，且在目标、预期、术语等方面[84, 94]存在差异，所以相互理解[94]、协同意识和利益均衡[89]尤为重要，需要以化解冲突为目标建立常态化沟通协商机制、重要工作协同推进机制、重点项目双向参与机制、内外部监督评价机制等保障机制[13, 84, 89, 94]。最后，需要关注外部环境，注重产学研用深度合作，构建共生演化的良性协同治理生态[29]。

除上述内容外，实现主体协同还需要厘清职责边界。第一种情况是跨机构的主体协同，即档案机构（包括档案局和档案馆）和大数据管理机构的协同，可基于档案管理与数据管理的差异性厘清职责边界。数据管理侧重数据共享利用等现实需求，具有应用导向，关注数据内容及所含信息的现行和经济价值[11, 14]。档案管理侧重高价值信息的鉴定与保存[14]，除数据内容外，还关注承载它们的“容器”（如载体、格式等），注重维护数据的原始性及凭证/证据与文化/记忆/历史等方面的价值[14, 49, 95-96]。因此，档案局和档案馆需要将管理重心置于业务数据和文化数据治理[93]。第二种情况是机构内的主体协同，即组织机构内设的档案部门（如档案室）和数据部门（如数据中心）的协同。考虑到档案管理侧重非结构化数据、数据管理侧重结构化数据的现实[77, 97]，可将文档与内容管理作为立足点，将档案部门定位为非结构化数据、历史数据和凭证性数据的管理部门，将档案管理系统定位为非结构化数据管理平台，将档案工作者定位为“内容管事”[84, 93]。

3.3.3 制度协同

制度协同涉及法律法规和制度规范两个层面。在法律法规层面，王玉珏等[11, 92]和陈怡[98]基于我国立法现实，对比分析《网络安全法》《数据安全法》《个人信息保护法》等数据法律与《档案法》等档案法律在规制对象、责任主体、管理环节等方面的异同，发现二者在归档、长期保存、分级分类、跨境流动、数据保护等方面存在衔接不畅的问题，并提出针对性对策。在数据保护立法备受重视的时代，也有学者开始关注被遗忘权等特定数据权利与档案法律

及实践的衔接问题[99-102]。

在制度规范层面，统一的政策、策略和程序是避免冲突的必要举措[95]，档案和数据制度规范需要强化相互衔接、参照和映射，为解决共性问题提供具体方法[87]，并将业务协同（3.3.4）要求融入制度设计中。从内容层面来看，制度协同包括政策合作、政策协调和政策整合三个阶段，政策合作侧重单项制度标准的合作，解决特定业务域的共性需求；政策协调侧重多项制度之间的关联，关注跨部门制度的内在一致性问题；政策整合旨在产出一体化制度体系[84]。从形式层面来看，制度协同的理想状态是内容层面的相互包含，次之是形式上的相互引用，如果上述两点均无法实现，制度制定主体至少需要提供档案和数据制度规范如何保持一致的指导说明[61]。具体到科技领域，科研档案和科学数据需要在归档管理（含鉴定原则、归档范围、元数据及格式要求等）、长期保存、数据安全、监督指导、开放审核等方面建立协同制度体系[103]。

3.3.4 业务协同

业务协同是指站在全局视角管理档案和数据，整合被视为不同领域的档案管理和数据管理，采取措施融合二者所用工具和技术[77]，在分析异同、找准协同点的基础上实现共赢[13]。

首先是收集阶段的协同。在收集前，由于档案部门和数据部门的数据均来源于数据形成部门，所以需要“摸清家底”，对业务活动中形成的数据进行全面、细致的调查，协同开展数据调查工作[13]。在收集过程中，需重点考虑三个方面：一是收集范围的协同。档案部门和数据部门应联动梳理各自的资源管辖范围，尽可能避免数据资源的重复或流失[13,84]。二是收集方式的协同。对档案部门而言是物理归档的数据，对数据部门而言就是逻辑归档；反之亦然[84]。三是收集流程和要求的协同。从优化工作流的角度出发，档案部门和数据部门可将数据收集流程合并[42]，并在整体框架下共同进行数据质量管控，维护数据的可知、可用和可控[14]。在数据部门快速崛起的背景下，档案部门只有通过协同才能更好地履行接收者的角色，成为对国家和社会有保存价值的数据的最终归宿地。

其次是管理阶段的协同，主要是指不同数据组织方式的融合。一是分类协同。档案部门和数据部门可以共同构建基于职能的档案和数据一体化分类体系，建立覆盖全局、以用户为核心的数据地图[84]；或结合档案领域来源原则和数据领域主题事由等不同分类原则，对档案与数据进行统一多维分类[88]。二是著录（元数据）协同，即融合两个领域的著录方法，建立元数据的关联关系。在科研领域，档案部门和科学数据仓储在多级著录、多实体著录、受控词表的使用等方面存在共性[49,104-105]，能够为相互协作、资源整合奠定基础。例如，《信息技

术 数据溯源描述模型》(GB/T 34945-2017)中的数据、活动、执行主体等实体，与《信息与文献 文件管理流程 文件元数据》(ISO 23081-1：2017)中的文件、业务、责任者等实体基本一致。

再次是保存阶段的协同。一是长期保存协同。档案部门与数据部门需在长期保存目标、制度规范、策略与方法等方面实现优势互补[84]，在安全存储领域建立数据互备份机制，提升数据安全管理效能[42]。二是处置协同。双方需要分析档案保管期限表与数据更新/删除/退役周期的关系[77]，据此构建协调一致的处置标准，确保同一业务对象的档案保管期限等于或高于数据保管期限[84]。

4 讨论与展望

整体来看，当前研究呈现出两点规律：第一，从接收者到指导者再到协同者，相关文献的平均出版时间越来越近，在一定程度上反映出档案部门在参与数据管理过程中角色定位的演变发展，当然，现实情况往往是多种角色并存；第二，接收者是中外共同关注的话题，但相对而言，国外研究侧重指导者，国内研究侧重协同者。这些研究内容较为丰富，在接收者场景下，提出识别归档范围、制定归档思路的方式方法；在指导者场景下，通过理论推导和实证检验，论证了档案理论、方法与工具在数据管理（尤其是科学数据管理）中的适用性以及档案部门作为指导者参与数据管理的可能性；在协同者场景下，明确了档案部门与数据部门协同开展数据管理的动因与框架，并围绕主体、制度等协同要素展开深入探索。但仍然存在一定局限，需要进一步拓展完善。

4.1 接收者视角：结构化数据归档方案有待升级

关于接收者的相关研究，主要存在 3 点局限：一是结构化数据的归档对象仍较为模糊。采用数据形成和数据利用的不同视角可能产生关于归档对象的不同认知。当前研究提出的“档案数据体”[18]是与数据环境相适应且具有灵活性的概念，但究竟如何在数据驱动的业务系统中识别“档案数据体”、如何应用“档案数据体”这一概念，并不清晰。二是当前研究提出的物理归档方法普遍存在成本过高、语义缺失等问题，尤其是长期保存和方便利用之间存在突出矛盾，归档后的开发利用尚未得到解决[15]。例如，在澳大利亚国家档案馆早期探索中，由于缺乏技术能力和专业知识，没有为石油数据建立必要的智力控制（如著录等），从而无法对这些数据进行检索利用[106]。三是逻辑归档的实现机理和路径方法尚未明确。与世纪之交的集中与分布保管模式之争类似，不能仅仅停留在何种归档思路更科学合理的层面，

而是要结合实际情况明确在何种情况下采用何种归档思路。就逻辑归档而言，亟需解决基于元数据的统一目录、档案化管理要求、档案实体和管理权分离情况下档案部门的监督保障机制等。因此，需要进一步研究如何在高度结构化、集成化的数据环境中识别归档对象、界定档案构成要素，探究凭证性维护、语义保存和开发利用之间的平衡机制，在收集更多实证数据的基础上深入探索逻辑归档的应用理论与实践方案。

4.2 指导者视角：多场景下档案部门指导数据管理的实现机制及效果有待探索

关于指导者的相关研究，主要存在 2 点局限：一是虽然论证了档案部门作为指导者参与数据管理的潜在可能性和必要性，但并未深入探索如何实现及实现效果等应用层面的问题。理想与现实之间不可避免地存在落差。例如，Elizabeth Shepherd 等学者基于《国际档案著录规则（总则）》[ISAD(G)]，延续其多级著录的特点，重点对系列级、案卷级的著录项进行扩展和优化，形成目前已知、唯一的基于档案著录标准的结构化数据集著录标准[107]，并曾应用于隶属于英国国家档案馆的国家数据集数字档案馆（National Digital Archive of Datasets，NDAD）[33]，但 NDAD 已于 2010 年停用，且英国国家档案馆没有继续采用基于 ISAD(G)的数据集著录标准[104]。该案例表明档案理论、方法与工具应用于数据环境可能会面临各种现实问题、挑战和阻力。二是在多数指导者相关研究中，研究者面临（或预设）的场景是，数据形成部门面临数据管理的相关难题，在没有数据部门的情况下，档案部门以其积累的经验辅助形成部门进行数据管理，侧重形成部门和档案部门的交互。但现实中可能会存在数据部门、图书馆、IT 部门、信息中心等其他指导者角色，当前研究较少探讨不同的指导者角色之间的关系以及如何处理这种关系。因此，需要进行深入的识别和分析，研究档案部门如何指导数据管理活动，尤其是当数据管理涉及多种指导者角色时，需进一步探索如何处理档案部门的指导者角色与其他指导者角色的关系、如何实现不同指导者角色的合作与协调等问题。

4.3 协同者视角：多要素协同推进路径及协同意愿与动机研究有待深化

关于协同者的相关研究，主要存在 3 点局限：一是研究重心在主体和制度要素，资源、服务、技术等要素的协同受到的关注较少。档案部门作为协同者参与数据管理的研究仍处于起步阶段，但随着研究的深入，需要逐步探索其他要素的协同。二是业务协同的研究深度不够。业务要素在协同要素中居于核心位置，虽然当前研究探讨了业务协同的方向性和原则性内容，但尚未深入管理生命周期、质量管控、分类、元数据、存储与保存等方面的具体协同与融通策略。此外，档案资源与数据资源汇聚整合的方式方法、档案与数据融合应用的成效

及价值规律等内容也都有待深入的实证研究和理论提炼。三是缺乏利益相关者视角下协同意愿和动机研究。当前研究对协同动因的分析多站在国家宏观层面或档案部门的角度，缺少数据部门的视角，而实践中的协同需要多主体配合。研究表明，从数据部门的视角来看，协同能给数据管理带来什么并不明确，数据部门参与协同管理的动机和意愿不如档案部门[13]。因此，需要在主体、制度等要素协同的基础上，以业务协同为轴心，兼顾资源、服务、技术等其他要素，统筹考虑档案部门与数据部门的利益诉求，考察数据部门等其他利益相关者的的协同态度和意愿，为档案部门作为协同者参与数据管理提出更为务实的针对性策略。

5 结语

档案与数据、档案管理与数据管理有着天然联系。档案参与数据管理已经引起一些国家档案馆的关注。美国国家档案与文件署发布《机构首席数据官与文件管理项目的协调》，指出“以前认为文件管理和数据管理是不同领域……现在应将其视为相关领域，并在机构信息管理框架内进行协作、互惠和更紧密的整合”[108]。档案工作者和文件管理者在实际工作中促进和执行数据管理并非新鲜事，档案领域可以为数据管理提供有用的建议[109]。正如冯惠玲教授所言，“文件和档案管理是数据管理中具有特殊使命、特殊功能的重要方面军，我们必须理直气壮地以‘在其中’的姿态证明自身的意义和价值”[110]。未来需要结合档案部门的接收者、指导者、协同者等角色定位展开持续深入的研究，在数据社会发出档案声音、贡献档案力量。

参考文献

[1] 钱毅.技术变迁环境下档案对象管理空间演化初探[J].档案学通讯,2018(2):10-14.

[2] 何思源,袁焱艳,姜盛新等.我国数字转型相关规划中的“档案参与”分析及启示[J].档案与建设,2022(9):14-17.

[3] 刘越男,周文泓,李雪彤等.我国档案事业“十四五”发展图景展望——基于国家及省级地方档案事业“十四五”规划的文本分析[J/OL].图书情报知识:1-9[2023-02-10].http://gfgga60aabc7d15084b00hfoc0pxxqwwoq60uq.fhaz.libproxy.ruc.edu.cn/kcms/detail/42.1085.g2.20230202.1117.001.html.

[4] 高晨翔,牛力.国内“档案数据”研究述评[J].档案学研究,2020(5):11-18.

[5] 金波,周枫,杨鹏.档案数据研究进展与研究题域[J].情报科学,2021,39(11):187-193.

[6] 杨茜茜.概念、内容与定位：我国档案数据治理研究的理论坐标[J].档案学研究,2021(6):28-34.

[7] GRANT R. Recordkeeping and research data management: a review of perspectives[J].Records management journal, 2017, 27(2):159-174.

[8] DAMA 国际.DAMA 数据管理知识体系指南[M].DAMA 中国分会翻译组译.北京:机械工业出版社,2020:1,11.

[9] 李忱博.当代档案学代表性英文期刊研究[J].档案与建设,2014(1):23-27.

[10] 中华人民共和国档案法[EB/OL].[2023-02-07].http://www.gov.cn/xinwen/2020-06/21/content_5520875.htm.

[11] 王玉珏,吴一诺.档案法律融入数据法律体系的内在逻辑、问题与路径[J].档案学研究,2022(3):28-35.

[12] 周克华.现代企业的文件与数据[J].中国档案,2009(3):36-37.

[13] 戴柏清,秦顺,何思源.科研档案与科学数据利益主体协同机制探析[J].档案学通讯,2023(1):61-70.

[14] 梁凯.多维度视角下大数据与档案关系的思考——以杭州市大数据管理为背景[J].中国档案,2018(7):66-67.

[15] 李泽锋.基于数据仓库的数据库电子文件管理研究[J].档案学通讯,2007(5):65-68.

[16] 金更达,潘燕军.结构化数据长期保存问题探析[J].档案学通讯,2006(5):60-64.

[17] DINGWALL G, MARCIANO R, MOORE R, et al. From data to records: preserving the geographic information system of the city of Vancouver[J]. Archivaria, 2007, 64(1):181-198.

[18] 刘越男.大数据政策背景下政务文件归档面临的挑战[J].档案学研究,2018(2):107-114.

[19] 周林兴,邓晋芝.错位的价值判定应该被纠正——对档案鉴定及销毁的历史视域考量[J].档案学通讯,2015(2):38-42.

[20] 杨冬权.从“选时代”到“全时代”——智慧社会档案工作的历史性转折[J].中国档案,2021(1):76-80.

[21] 郑金月.数据价值：大数据时代档案价值的新发现[J].浙江档案,2015(12):11-14.

[22] 谢丽.档案鉴定与鉴定档案：去留之间超越信息技术及政府投入的考量[J].档案学通讯,2015(5):18-24.

[23] 孙大东.档案鉴定的历史和现实视域考量——与周林兴、邓晋芝二位作者商榷[J].档案学通讯,2015(5):13-17.

[24] YEO G. Can we keep everything? The future of appraisal in a world of digital profusion[M]// BROWN C. Archival futures. London: Facet Publishing, 2018:45-63.

[25] 陶水龙.三维数据归档策略研究[J].档案学研究,2018(6):101-104.

[26] MCDONALD J, LEVEILLE MV. Whither the retention schedule in the era of big data and open data?[J]. Records management journal, 2014, 24(2):99-121.

[27] MCDONALD J. Scheduling data in systems: three PAC pilot projects[J]. Archivaria, 1985, 20(1):241-244.

[28] 王强,高强.业务系统数据归档研究——以中国石油业务系统数据归档实践为例[J].浙江档案,2019(12):36-39.

[29] 刘越男.数据管理大潮下电子文件管理的挑战与对策[J].北京档案,2021(6):4-9.

[30] 钱毅.数据态环境中数字档案对象保存问题与策略分析[J].档案学通讯,2019(4):40-47.

[31] DARROCH G, GAVREL S. Preserving historical databases and facing technical change: common issues for social historians and archivists[J]. Archivaria, 1992, 34(1):288-297.

[32] LONG S M. Documenting federal scientific and technical information (STI): a discussion of appraisal criteria and applications for the National Archives and Records Administration[J]. Journal of government information, 1995, 22(4):311-319.

[33] SLEEMAN P. It's public knowledge: the National Digital Archive of Datasets[J]. Archivaria, 2004, 58(1):173-200.

[34] STRUYK R J, 丁怡, 李刚. 智库研究项目文件与数据的归档问题[J]. 图书馆论坛,2018,38(4):20-24.

[35] 金波.地理信息库数据切片归档实现方式研究[J].档案与建设,2017(2):32-36.

[36] 钱毅,刘力超.数据库电子文件归档与长期保存技术路径研究[J].档案学研究,2017(4):67-72.

[37] ABRAMS B. Database preservation case study: testing SIARD 2.0 at NARA[EB/OL]. [2022-10-30].https://www.dpconline.org/docs/miscellaneous/events/2020-events/2294-siard-dpc-presentation/file.

[38] National Archives of Korea. Record keeping criteria for dataset - composition of data set management reference table & exchange of dataset - version 1.0[EB/OL].[2022-10-30].https://contents.archives.go.kr/next/data/standardCondition.do.

[39] FRANCIS P, FOWLER D, MCCORMACK J, et.al. SIARD research (2013-0387) 201

4-15 Report[EB/OL].[2022-10-30].https://prov.vic.gov.au/sites/default/files/files/Govt%20Services%20General/SIARD%20Research%202014-15%20Report.pdf.
[40] 云霄.基于 ERP 系统结构化数据归档技术方案研究[J].北京档案,2014(7):34-36.
[41] 高闯,柳林集.合规与妥协：空客德国产品数据归档的现状及其启示[J].档案学研究,2021(2):119-124.
[42] 蔡盈芳.推进科学数据与科研档案的协同管理[J].中国档案,2021(9):60-61.
[43] 赵跃,孙晶琼,段先娥.档案化：档案科学介入数据资源管理的理性思考[J].档案学研究,2020(5):83-91.
[44] 杨茜茜.智慧社会背景下政务服务数据资源的档案化管理——基于电子证照应用管理的研究[J].浙江档案,2018(9):10-12.
[45] LEE C A, TIBBO H. Where's the archivist in digital curation? exploring the possibilities through a matrix of knowledge and skills[J]. Archivaria, 2011, 72(1):123-168.
[46] ALTMAN M, ADAMS M O, CRABTREE J, et.al. Digital preservation through archival collaboration: the data preservation alliance for the social sciences[J]. American archivist, 2009, 72(1):170-184.
[47] AKMON D, ZIMMERMAN A, DANIELS M, et al. The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs[J]. Archival science, 2011, 11(3-4):329-348.
[48] FEAR K, DONALDSON D. Provenance and credibility in scientific data repositories[J]. Archival science, 2012, 12(3):319-339.
[49] YARMEY K A, YARMEY L R. All in the family: A dinner table conversation about libraries, archives, data, and science[J/OL].Archive journal[2022-12-05].http://www.archivejournal.net/essays/all-in-the-family-a-dinner-table-conversation-about-libraries-archives-data-and-science/.
[50] NOONAN D, CHUTE T. Data curation and the university archives[J]. American archivist, 2014, 77(1):201-240.
[51] EVANS J, REED B, LINGER H, et al. Winds of change: a recordkeeping informatics approach to information management needs in data-driven research environments[J]. Records management journal, 2014, 24(3):205-223.
[52] CHILDS S, MCLEOD J, LOMAS E, et al. Opening research data: issues and opportunities[J].

Records management journal, 2014, 24(2):142-162.

[53] POOLE A H. How has your science data grown? digital curation and the human factor: a critical literature review[J]. Archival science, 2015, 15(2):101-139.

[54] FRANK R D, YAKEL E, FANIEL I M. Destruction/reconstruction: preservation of archaeological and zoological research data[J]. Archival science, 2015, 15(2):141-167.

[55] 毛天宇.数字监护研究中档案学理论的应用及启示探析[J].档案学通讯,2016(1):34-38.

[56] BORGERUD C, BORGLUND E. Open research data, an archival challenge?[J]. Archival science, 2020, 20(3):279-302.

[57] 王宁,刘越男.档案学视角下的科学数据管理——基于国际组织相关成果的研究[J].图书情报工作,2021,65(5):88-97.

[58] 于英香,张雅颉.“档案参与”科学数据监管：缘起、现状与动因[J].档案学研究,2021(2):104-110.

[59] MCLEOD J. Thoughts on the opportunities for records professionals of the open access, open data agenda[J]. Records management journal, 2012, 22(2):92-97.

[60] SERRA L E C. The mapping, selecting and opening of data: the records management contribution to the open data project in Girona city council[J]. Records management journal, 2014, 24(2):87-98.

[61] LEVEILLE V, TIMMS K. Through a records management lens: creating a framework for trust in open government and open government information[J].Canadian journal of information and library science, 2015, 39(2):154-190.

[62] 周文泓.开放数据背景下的文档管理探析[J].北京档案,2018(8):11-14.

[63] SHEPHERD E, BUNN J, FLINN A, et al. Open government data: critical information management perspectives[J]. Records management journal, 2019, 29(1-2):152-167.

[64] LEMIEUX V L, GORMLY B, ROWLEDGE L. Meeting big data challenges with visual analytics: the role of records management[J]. Records management journal, 2014, 24(2):122-141.

[65] DALE K L. RIM’s role in harnessing the power of dig data[J]. Information management, 2015, 49(4):29-32.

[66] SØDRING T, REINHOLDTSEN P, MASSEY D. A record-keeping approach to managing IoT-data for government agencies[J]. Records management journal, 2020, 30(2):221-239.

[67] 周林兴,林腾虹.基于文件连续体理论的政务数据治理体系优化研究[J].电子政

务,2021(4):114-124.

[68] 章燕华.数据档案化治理研究：一个研究分析框架[J].浙江档案,2022(11):25-29.

[69] 宋文凤.科学数据价值鉴定研究[D].长春:吉林大学,2013:8-9,24-30.

[70] WALLIS J C, BORGMAN C L, MAYERNIK M S, et.al. Moving archival practices upstream: an exploration of the life cycle of ecological sensing data in collaborative field research[J]. International journal of digital curation, 2008, 3(1):114-125.

[71] ESANU J, DAVIDSON J, ROSS S, et.al. Selection, appraisal, and retention of digital scientific data: highlights of an ERPANET/CODATA Workshop[J]. Data science journal, 2004, 3(2):226-232.

[72] MADAY C, MOYSAN M. Records management for scientific data[J]. Archives and manuscripts, 2014, 42(2):190-192.

[73] THURSTON A C. Trustworthy records and open data[J/OL].The journal of community informatics[2023-01-31].https://ojs.uwaterloo.ca/index.php/JoCI/article/view/3047/3951.

[74] CHORLEY K M. The challenges presented to records management by open government data in the public sector in England[J]. Records management journal, 2017, 27(2):149-158.

[75] 代佳欣.基于过程的政府数据开放风险识别与防控策略研究[J].情报杂志,2019,38(6):145-151.

[76] LARSON E. Big questions: digital preservation of big data in government[J]. American archivist, 2020, 83(1):5-20.

[77] MCDONALD J. Records management and data management: closing the gap[J]. Records management journal, 2010, 20(1):53-60.

[78] 何思源,刘越男,祁天娇.档案与数据协同管理的历史溯源与现实动因[J].北京档案,2023(1):8-12.

[79] BORGLUND E, ENGVALL T. Open data? data, information, document or record?[J]. Records management journal, 2014, 24(2):163-180.

[80] ANDERSON K. The footprint and the stepping foot: archival records, evidence, and time[J]. Archival science, 2013, 13(4):349-371.

[81] 何嘉荪,谭建月.档案概念再认识——大数据引起的思考[J].档案与建设,2017(8):4-6+10.

[82] 何思源.科研档案和科学数据协同治理的动因分析[J].档案管理,2020(6):64-66.

[83] 蔡盈芳.数据管理业务与档案融合管理研究[J].档案学研究,2021(3):40-46.

[84] 刘越男,何思源,王强等.企业档案与数据资产的协同管理：问题与对策[J].档案学研

究,2022(6):94-102.

[85] DOORN P, TJALSMA H. Introduction: archiving research data[J]. Archival science, 2007, 7(1):1-20.

[86] NIELSEN P. Merging cultures: Danish integration of academic data service into traditional archive system[J]. IASSIST quarterly, 1995, 19(2):27-35.

[87] 何思源,刘越男.科学数据和科研档案的管理协同：框架和路径[J].档案学通讯,2021(1):49-57.

[88] 刘越男,何思源.科学数据与科研档案的管理协同：调查与思考[J].图书情报工作,2022,66(1):96-105.

[89] 龚慧卿.地方大数据管理部门与档案主管部门协同发展研究[D].湘潭:湘潭大学,2021:27-32,37-40,45-46.

[90] 徐拥军,张臻,任琼辉.国家大数据战略背景下档案部门与数据管理部门的职能关系[J].图书情报工作,2019,63(18):5-13.

[91] 符京生,刘汉青,苏兴华等.大型企业档案与数据协同治理框架与实现路径[J].浙江档案,2020(12):56-57.

[92] 王玉珏,吴一诺,凌敏菡.《数据安全法》与《档案法》协调研究[J].图书情报工作,2021,65(22):24-34.

[93] 刘越男.数据治理：大数据时代档案管理的新视角和新职能[J].档案学研究,2020(5):50-57.

[94] RICHARDS L L. Teaching data creators how to develop an OAIS-compliant digital curation system: colearning and breakdowns in support of requirements analysis[J]. American archivist, 2016, 79(2):145-165.

[95] YEO G. Records, information and data: exploring the role of record-keeping in an information culture[M].London: Facet Publishing, 2018:199,193.

[96] 陈艳.后机构改革时代：档案机构定位的重塑——基于山东省省级机构档案工作者的调查[J].档案学研究,2020(3):39-45.

[97] 徐拥军,王露露.档案部门参与大数据战略的必备条件和关键问题——以浙江省为例[J].浙江档案,2018(11):11-14.

[98] 陈怡.新《档案法》与数据法律衔接研究[D].北京:中国人民大学,2022.

[99] XIE S L. Retention in “the right to be forgotten” scenario: a records management examination[J]. Records management journal, 2016, 26(3):279-292.

[100] VAVRA A N. The right to be forgotten: an archival perspective[J].American archivist, 2018, 81(1):100-111.

[101] 连志英.被遗忘权对图书馆档案馆信息存档及信息获取的影响[J].图书情报工作,2021,65(16):35-41.

[102] 龙家庆.被遗忘权对档案工作的影响[J].档案学研究,2022(1):43-50.

[103] 霍倩,梁宵萌,潘亚男.科学数据管理政策对档案部门科学数据管理的启示——基于政策工具的文本量化分析[J].图书情报工作,2021,65(20):23-30.

[104] NIU J. Organisation and description of datasets[J]. Archives and manuscripts, 2016, 44(2):73-85.

[105] NIU J. Aggregate control of scientific data[J]. Archives and records, 2016, 37(1):53-64.

[106] YORKE S. Management of petroleum data records in the custody of Australian archives[J]. Archives and manuscripts, 1997, 25(1):62-73.

[107] SHEPHERD E, SMITH C. The application of ISAD(G) to the description of archival datasets[J]. Journal of the society of archivists, 2000, 21(1):55-86.

[108] National Archives and Records Administration. Agency chief data officer coordination with records management programs[R/OL].[2022-11-15].https://www.archives.gov/files/records-mgmt/resources/cdo-rm-assessment-report.pdf.

[109] ROBERTS D. Defining electronic records, documents and data[J]. Archives and manuscripts, 1994, 22(1):14-26.

[110] 冯惠玲.融入数据管理 做电子文件管理追风人[J].北京档案,2020(12):6-7.

A Review of Research Related to Records and Archives Departments Engaging with Data Management

He Siyuan[1,2]

[1] School of Information Resource Management, Renmin University of China, Beijing 100872

[2] Electronic Records Management Research Center, Renmin University of China, Beijing 100872

Abstract: [**Purpose/significance**] This paper explores the role of records and archives (R&A) departments in data management and the approaches for R&A departments to engage with data management in order to provide a solid theoretical foundation for R&A departments to improve their performance and make a meaningful contribution to the data society. [**Method/process**] Using systematic review methodology, this paper selected and analyzed 104 literature on R&A

departments engaging with data management. [**Result/conclusion**] The results of the analysis revealed that R&A departments have three distinct roles in data management: they serve as receiver, guide and integrator. As a receiver, the R&A department connects the data lifecycle with the records lifecycle through data filing. As a guide, the R&A department uses archival theories and methods such as provenance principle, appraisal, continuity management, long-term preservation, and archival control tools such as classification schemes and catalogues, to provide advice on scientific data management, open government data, and big data applications. As a integrator, the R&A department collaborates with data departments, considering factors such as policies, organizations, operations, resources, services, and systems. Further research requires more in-depth theoretical exploration and empirical studies on the identification of filing objects and the upgrad of filing solutions for structured data, the implementation mechanisms and effectiveness evaluation of R&A departments' guidance on data management, the business integration between R&A departments and data departments including the motivations and willingness for integration management.

Keywords: records and archives management　data management　data governance　data archivalization